

Secrecy results for compound wiretap channels

Igor Bjelaković, Holger Boche, and Jochen Sommerfeld

Lehrstuhl für theoretische Informationstechnik, Technische Universität München, 80290 München, Germany
Email: {igor.bjelakovic, boche, jochen.sommerfeld}@tum.de

Abstract—We derive a lower bound on the secrecy capacity of the compound wiretap channel with channel state information at the transmitter which matches the general upper bound on the secrecy capacity of general compound wiretap channels given by Liang et al. and thus establishing a full coding theorem in this case. We achieve this with a stronger secrecy criterion and with a decoder that is robust against the effect of randomisation in the encoding. This relieves us from the need of decoding the randomisation parameter which is in general not possible within this model. Moreover we prove a lower bound on the secrecy capacity of the compound wiretap channel without channel state information and derive a multi-letter expression for the capacity in this communication scenario.

I. INTRODUCTION

Compound wiretap channels are among the simplest non-trivial models incorporating the requirement of security against a potential eavesdropper while at the same time the legitimate users suffer from channel uncertainty. They may be considered therefore as a starting point for theoretical investigation tending towards applications, for example, in wireless systems, a fact explaining an alive research activity in this area in recent years (cf. [1], [2] and references therein).

In this paper we consider finite families of pairs of channels $\mathfrak{W} = \{(W_t, V_t) : t = 1, \dots, T\}$ with common input alphabet and possibly different output alphabets. The legitimate users control W_t and the eavesdropper observes the output of V_t . We will be dealing with two communication scenarios. In the first one only the transmitter is informed about the index t (channel state information (CSI) at the transmitter) while in the second the legitimate users have no information about that index at all (no CSI). This setup is a generalisation of Wyner's [3] wiretap channel.

Along the way we will comment what our results look like when applied to widely used class of models of the form $\mathfrak{W} = \{(W_t, V_s) : t = 1, \dots, T, s = 1, \dots, S\}$ with $T \neq S$ which are special cases of the model we are dealing with in this paper.

Our contributions are summarised as follows: In [1] a general upper bound on the capacity of compound wiretap channel as the minimum secrecy capacity of the involved wiretap channels was given. We prove in Section III-B that the models whose secrecy capacity matches this upper bound contain all compound wiretap channels with CSI at the transmitter. At the same time we achieve this bound with a substantially stronger security criterion employed already in [4], [5], [6], and [7]. Indeed, our security proof follows closely that developed in [7] for single wiretap channel with classical input and quantum

output. In order to achieve secrecy we follow the common approach according to which randomised encoding is a permissible operation. The impact of randomisation at the legitimate decoder's site is usually compensated by communicating to her/him the outcome of the random experiment performed. However, in the case of compound wiretap channel with CSI at the transmitter this strategy does not work as is illustrated by an example in Section IV-A. We resolve this difficulty by developing a decoding strategy which is independent of the particular channel realisation and is insensitive to randomisation while decoding just at the optimal secrecy rate for all channels $\{W_t : t = 1, \dots, T\}$ simultaneously.

Moreover, a slight modification of our proofs allows us to determine the capacity of the compound wiretap channel without CSI by a (non-computable) multi-letter expression. This is content of Section III-C. We should mention, however, that the traditional proof strategy of sending the pair consisting of message and randomisation parameter to the legitimate receiver works as well in the case where the transmitter has no CSI.

In Section IV-B we give an example of compound wiretap channel such that both the set of channels to the legitimate receiver and to the eavesdropper are convex but whose secrecy capacities with CSI and without CSI at the transmitter are different. Indeed the former is positive while the latter is equal to 0.

Section III-D is devoted to the practically important model $\mathfrak{W} = \{(W_t, V_s) : t = 1, \dots, T, s = 1, \dots, S\}$ with the assumption that the transmitter has CSI for the T -part but has no CSI for the S -part of the channel. Here again we provide a multi-letter expression for the capacity. Additionally, we give a computable description of the secrecy capacity in the case where the channels to the eavesdropper are degraded versions of those to the legitimate receiver.

Our results are easily extended to arbitrary sets (even uncountable) of wiretap channels via standard approximation techniques [8].

II. COMPOUND WIRETAP CHANNEL

A. Definitions

Let A, B, C be finite sets and $\theta = \{1, \dots, T\}$ an index set. We consider two families of channels $W_t : A \rightarrow \mathcal{P}(B)^1$, $V_t : A \rightarrow \mathcal{P}(C)$, $t \in \theta$, which we collectively abbreviate by \mathfrak{W} and call the compound wiretap channel generated by the given families of channels. Here the first family represents the communication link to the legitimate receiver while the output

¹ $\mathcal{P}(B)$ denotes the set of probability distributions on B .

of the latter is under control of the eavesdropper. In the rest of the paper expressions like $W_t^{\otimes n}$ or $V_t^{\otimes n}$ stand for the n -th memoryless extension of the stochastic matrices W_t , V_t .

An (n, J_n) code for the compound wiretap channel \mathfrak{W} consists of a stochastic encoder $E : \mathcal{J}_n \rightarrow \mathcal{P}(A^n)$ (a stochastic matrix) with a message set $\mathcal{J}_n := \{1, \dots, J_n\}$ and a collection of mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. The maximum error probability of a (n, J_n) code \mathcal{C}_n is given by

$$e(\mathcal{C}_n) := \max_{t \in \theta} \max_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E(x^n | j) W_t^{\otimes n}(D_j^c | x^n). \quad (1)$$

I.e. neither the sender nor the receiver have CSI.

If channel state information is available at the transmitter the notion of (n, J_n) code is modified in that the encoding may depend on the channel index while the decoding sets remain universal, i.e. independent of the channel index t . The probability of error in (1) changes to

$$e_{\text{CSI}}(\mathcal{C}_n) := \max_{t \in \theta} \max_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E_t(x^n | j) W_t^{\otimes n}(D_j^c | x^n).$$

We assume throughout the paper that the eavesdropper always knows which channel is in use.

Definition 2.1: A non-negative number R is an achievable secrecy rate for the compound wiretap channel \mathfrak{W} with or without CSI respectively if there is a sequence $(\mathcal{C}_n)_{n \in \mathbb{N}}$ of (n, J_n) codes such that

$$\lim_{n \rightarrow \infty} e(\mathcal{C}_n) = 0 \text{ resp. } \lim_{n \rightarrow \infty} e_{\text{CSI}}(\mathcal{C}_n) = 0,$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log J_n \geq R,$$

and

$$\lim_{n \rightarrow \infty} \max_{t \in \theta} I(J; Z_t^n) = 0, \quad (2)$$

where J is a uniformly distributed random variable taking values in \mathcal{J}_n and Z_t^n are the resulting random variables at the output of eavesdropper's channel $V_t^{\otimes n}$.

The secrecy capacity in either scenario is given by the largest achievable secrecy rate and is denoted by $C_S(\mathfrak{W})$ and $C_{S, \text{CSI}}(\mathfrak{W})$.

B. Hints on operational meaning of strong secrecy

A weaker and widely used security criterion is obtained if we replace (2) by $\lim_{n \rightarrow \infty} \max_{t \in \theta} \frac{1}{n} I(J; Z_t^n) = 0$. We prefer to follow [4], [6], and [7] and require the validity of (2). A nice discussion on interrelation of several secrecy criteria is contained in [2]. We confine ourselves to giving some hints on the operational meaning of the requirement (2). To this end we restrict our attention to the case where the transmitter has no CSI in order to simplify our notation. The case of compound wiretap channel with CSI at the transmitter can be treated accordingly. Set

$$\varepsilon_n := \max_{t \in \theta} I(J; Z_t^n) \text{ with } \lim_{n \rightarrow \infty} \varepsilon_n = 0.$$

Then Pinsker's inequality implies that

$$\|p_{JZ_t^n} - p_J \otimes p_{Z_t^n}\| \leq c\sqrt{\varepsilon_n} \quad \forall t \in \theta, \quad (3)$$

with a positive universal constant c , where $\|\cdot\|$ is the variational distance. Suppose that the eavesdropper chooses for each $t \in \theta$ decoding sets $\{K_{j,t} \subset C^n : j \in \mathcal{J}_n\}$ with $C^n = \bigcup_{j \in \mathcal{J}_n} K_{j,t}$. We will lower bound the average error probability (and consequently the maximum error probability) for every choice of the decoding rule the eavesdropper might make. Set

$$e_{\text{av}}(t) := \frac{1}{J_n} \sum_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E(x^n | j) V_t^{\otimes n}(K_{j,t}^c | x^n).$$

Then

$$\begin{aligned} e_{\text{av}}(t) &= \sum_{j \in \mathcal{J}_n} p_{JZ_t^n}(\{j\} \times K_{j,t}^c) \\ &= p_{JZ_t^n} \left(\bigcup_{j \in \mathcal{J}_n} \{j\} \times K_{j,t}^c \right) \\ &\geq p_J \otimes p_{Z_t^n} \left(\bigcup_{j \in \mathcal{J}_n} \{j\} \times K_{j,t}^c \right) - c\sqrt{\varepsilon_n} \\ &= \sum_{j \in \mathcal{J}_n} p_J \otimes p_{Z_t^n}(\{j\} \times K_{j,t}^c) - c\sqrt{\varepsilon_n} \\ &= \frac{1}{J_n} \sum_{j \in \mathcal{J}_n} p_{Z_t^n}(K_{j,t}^c) - c\sqrt{\varepsilon_n} \\ &= \frac{J_n - 1}{J_n} - c\sqrt{\varepsilon_n} \\ &= 1 - \frac{1}{J_n} - c\sqrt{\varepsilon_n}, \end{aligned} \quad (4)$$

where in the second and the fourth line we have used the fact that the sets $\{j\} \times K_{j,t}^c$, $j \in \mathcal{J}_n$, are mutually disjoint, the third line follows from (3), and in the sixth line we merely observed that for any non-negative numbers a_1, \dots, a_J with $\sum_{j=1}^J a_j = 1$ we have $\sum_{j=1}^J (1 - a_j) = J - 1$. Consequently, the average (and hence maximum) error probability of every decoding strategy the eavesdropper might select tends to 1 as soon as $J_n \rightarrow \infty$. It should be remarked, however, that although for the vast majority of messages the eavesdropper will be in error there is still a possibility left that she/he can decode a small fraction of them correctly.

As will follow from the proofs below we will have $\varepsilon_n = 2^{-na}$, $a > 0$, and $J_n = 2^{nR}$, $R > 0$, if the secrecy capacity is positive so that the speed of convergence in (4) will be exponential.

Notice that (3) means that the random variables Z_t^n at the output of the channel to the eavesdropper are almost independent of the random variable J embodying the messages to be transmitted to the legitimate receiver. Therefore it is heuristically convincing that our criterion (2) offers secrecy to some extent for communication tasks going beyond the transmission of messages. To demonstrate this by an example we introduce, based on [9], the notion of identification attack as follows. Suppose that for each fixed $t \in \theta$ and any $j \in \mathcal{J}_n$ there is a subset $K_{j,t} \subset C^n$ on the eavesdropper's output alphabet where now the sets $K_{j,t}$ need not necessarily be mutually disjoint. With $E : \mathcal{J}_n \rightarrow \mathcal{P}(A^n)$ being the stochastic encoder used to transmit (!) messages to the legitimate receiver

we can write down the identification errors of first and second kind (cf. [9] for further explanation of this code concept) for the eavesdropper's channel as

$$\sum_{x^n \in A^n} E(x^n | j) V_t^{\otimes n}(K_{j,t}^c | x^n), \quad (5)$$

and

$$\sum_{x^n \in A^n} E(x^n | i) V_t^{\otimes n}(K_{j,t} | x^n) \quad (6)$$

for $j, i \in \mathcal{J}_n$, $i \neq j$.

One possible interpretation of this attack, again based on [9], is that on the eavesdropper's side of the channel there are persons F_1, \dots, F_{J_n} observing the output of the channel. The sole interest of F_j is whether or not the message j has been sent to the legitimate receiver. Thus F_j performs the hypothesis test represented by $K_{j,t}$ based on his/her knowledge of $t \in \theta$ and (5), (6) are just the errors of the first resp. second kind for that hypothesis test.

Let us define for $j \in \mathcal{J}_n$

$$g(j, t) := \sum_{x^n \in A^n} \left(E(x^n | j) V_t^{\otimes n}(K_{j,t}^c | x^n) + \frac{1}{J_n - 1} \sum_{\substack{i=1 \\ i \neq j}}^{J_n} E(x^n | i) V_t^{\otimes n}(K_{j,t} | x^n) \right),$$

which is a number in $[0, 2]$.

Notice that if

$$g(j, t) \geq 1 - \eta$$

for some $\eta \in (0, 1)$ then either

$$\sum_{x^n \in A^n} E(x^n | j) V_t^{\otimes n}(K_{j,t}^c | x^n) \geq \frac{1 - \eta}{2},$$

or there is at least one $i \neq j$ with

$$\sum_{x^n \in A^n} E(x^n | i) V_t^{\otimes n}(K_{j,t} | x^n) \geq \frac{1 - \eta}{2},$$

or both, so that no reliable identification of message j can be guaranteed. We show now that under assumption of (2) we have

$$\frac{1}{J_n} \sum_{j=1}^{J_n} g(j, t) \geq 1 - \eta_n, \quad \eta_n = o(n^0) \quad (7)$$

so that at most a fraction $\frac{2}{3}(1 + \eta_n)$ of $j \in \mathcal{J}_n$ can satisfy the inequality

$$g(j, t) < \frac{1}{2}.$$

This last assertion is readily seen from (7) by applying Markov's inequality to the set

$$F := \{j \in \mathcal{J}_n : 2 - g(j, t) > \frac{3}{2}\}.$$

In order to prove (7), note that for any $t \in \theta$

$$\begin{aligned} & \frac{1}{J_n} \sum_{j=1}^{J_n} g(j, t) \\ &= \sum_{j=1}^{J_n} \left(p_{JZ_t^n}(\{j\} \times K_{j,t}^c) + \frac{1}{J_n - 1} p_{JZ_t^n}(\{j\}^c \times K_{j,t}) \right) \\ &= p_{JZ_t^n} \left(\bigcup_{j \in \mathcal{J}_n} \{j\} \times K_{j,t}^c \right) + \frac{1}{J_n - 1} \sum_{j=1}^{J_n} p_{JZ_t^n}(\{j\}^c \times K_{j,t}) \\ &\geq p_J \otimes p_{Z_t^n} \left(\bigcup_{j \in \mathcal{J}_n} \{j\} \times K_{j,t}^c \right) + \\ &+ \frac{1}{J_n - 1} \sum_{j=1}^{J_n} p_J \otimes p_{Z_t^n}(\{j\}^c \times K_{j,t}) - c\sqrt{\varepsilon_n} - c \frac{J_n}{J_n - 1} \sqrt{\varepsilon_n} \\ &= \frac{1}{J_n} \sum_{j=1}^{J_n} (p_{Z_t^n}(K_{j,t}^c) + p_{Z_t^n}(K_{j,t})) - c\sqrt{\varepsilon_n} \frac{2J_n - 1}{J_n - 1} \\ &= 1 - c\sqrt{\varepsilon_n} \frac{2J_n - 1}{J_n - 1}, \end{aligned}$$

where in the third line we have used (3) and in the fourth we inserted $p_J(\{j\}^c) = \frac{J_n - 1}{J_n}$.

Besides the attempts of the eavesdropper to decode or identify messages we can introduce attacks corresponding to each communication task introduced in [10]. It would be interesting, not only from the mathematical point of view, to see against which of them and to what extent secrecy can be guaranteed by the condition (2).

III. CAPACITY RESULTS

A. Preliminaries

In what follows we use the notation as well as some properties of *typical* and *conditionally typical* sequences from [11]. For $p \in \mathcal{P}(A)$, $W : A \rightarrow \mathcal{P}(B)$, $x^n \in A^n$, and $\delta > 0$ we denote by $\mathcal{T}_{p,\delta}^n$ the set of typical sequences and by $\mathcal{T}_{W,\delta}^n(x^n)$ the set of conditionally typical sequences given x^n in the sense of [11].

The basic properties of these sets that are needed in the sequel are summarised in the following three lemmata.

Lemma 3.1: Fixing $\delta > 0$, for every $p \in \mathcal{P}(A)$ and $W : A \rightarrow \mathcal{P}(B)$ we have

$$\begin{aligned} p^{\otimes n}(\mathcal{T}_{p,\delta}^n) &\geq 1 - (n+1)^{|A|} 2^{-nc\delta^2} \\ W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n) | x^n) &\geq 1 - (n+1)^{|A||B|} 2^{-nc\delta^2} \end{aligned}$$

for all $x^n \in A^n$ with $c = 1/(2 \ln 2)$. In particular, there is $n_0 \in \mathbb{N}$ such that for each $\delta > 0$ and $p \in \mathcal{P}(A)$, $W : A \rightarrow \mathcal{P}(B)$

$$\begin{aligned} p^{\otimes n}(\mathcal{T}_{p,\delta}^n) &\geq 1 - 2^{-nc'\delta^2} \\ W^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n) | x^n) &\geq 1 - 2^{-nc'\delta^2} \end{aligned}$$

holds with $c' = \frac{c}{2}$.

Proof: Standard Bernstein-Sanov trick using the properties of types from [11] and Pinsker's inequality. The details can be found in [12] and references therein for example. ■

Recall that for $p \in \mathcal{P}(A)$ and $W : A \rightarrow \mathcal{P}(B)$, $pW \in \mathcal{P}(B)$ denotes the output distribution generated by p and W and that $x^n \in \mathcal{T}_{p,\delta}^n$ and $y^n \in \mathcal{T}_{W,\delta}^n(x^n)$ imply that $y^n \in \mathcal{T}_{pW,2|A|\delta}^n$.

Lemma 3.2: Let $x^n \in \mathcal{T}_{p,\delta}^n$, then for $V : A \rightarrow \mathcal{P}(C)$

$$\begin{aligned} |\mathcal{T}_{pV,2|A|\delta}^n| &\leq \alpha^{-1} \\ V^n(z^n|x^n) &\leq \beta \quad \text{for all } z^n \in \mathcal{T}_{V,\delta}^n(x^n) \end{aligned}$$

hold where

$$\alpha = 2^{-n(H(pV)+f_1(\delta))} \quad (8)$$

$$\beta = 2^{-n(H(V|p)-f_2(\delta))} \quad (9)$$

with universal $f_1(\delta), f_2(\delta) > 0$ satisfying $\lim_{\delta \rightarrow \infty} f_1(\delta) = 0 = \lim_{\delta \rightarrow \infty} f_2(\delta)$.

Proof: Cf. [11]. ■

In addition we need a further lemma which will be used to determine the rates at which reliable transmission to the legitimate receiver is possible.

Lemma 3.3: Let $p, \tilde{p} \in \mathcal{P}(A)$ and two stochastic matrices $W, \tilde{W} : A \rightarrow \mathcal{P}(B)$ be given. Further let $q, \tilde{q} \in \mathcal{P}(B)$ be the output distributions, the former generated by p and W and the latter by \tilde{p} and \tilde{W} . Fix $\delta \in (0, \frac{1}{4|A||B|})$. Then for every $n \in \mathbb{N}$

$$q^{\otimes n}(\mathcal{T}_{\tilde{W},\delta}^n(\tilde{x}^n)) \leq (n+1)^{|A||B|} 2^{-n(I(\tilde{p}, \tilde{W})-f(\delta))}$$

for all $\tilde{x}^n \in \mathcal{T}_{\tilde{p},\delta}^n$ and

$$q^{\otimes n}(\mathcal{T}_{W,\delta}^n(x^n)) \leq (n+1)^{|A||B|} 2^{-n(I(p, W)-f(\delta))}$$

for all $x^n \in \mathcal{T}_{p,\delta}^n$ holds for a universal $f(\delta) > 0$ and $\lim_{\delta \rightarrow 0} f(\delta) = 0$.

Proof: Cf. [12]. ■

The last lemma is a standard result from large deviation theory.

Lemma 3.4: (Chernoff bounds) Let Z_1, \dots, Z_L be i.i.d. random variables with values in $[0, 1]$ and expectation $\mathbb{E}Z_i = \mu$, and $0 < \epsilon < \frac{1}{2}$. Then it follows that

$$\Pr \left\{ \frac{1}{L} \sum_{i=1}^L Z_i \notin [(1 \pm \epsilon)\mu] \right\} \leq 2 \exp \left(-L \cdot \frac{\epsilon^2 \mu}{3} \right),$$

where $[(1 \pm \epsilon)\mu]$ denotes the interval $[(1 - \epsilon)\mu, (1 + \epsilon)\mu]$.

B. CSI at the transmitter

The main result in this section is the following theorem.

Theorem 3.5: The secrecy capacity of the compound wiretap channel \mathfrak{W} with CSI at the transmitter is given by

$$C_{S,CSI}(\mathfrak{W}) = \min_{t \in \theta} \max_{V \rightarrow X \rightarrow (YZ)_t} (I(V, Y_t) - I(V, Z_t)).$$

Notice first that the inequality

$$C_{S,CSI}(\mathfrak{W}) \leq \min_{t \in \theta} \max_{V \rightarrow X \rightarrow (YZ)_t} (I(V, Y_t) - I(V, Z_t))$$

is trivially true since we cannot exceed the secrecy capacity of the worst wiretap channel in the family \mathfrak{W} . This has been already pointed out in [1]. The rest of this section is devoted to the proof of the achievability.

Proof: We choose $p_1, \dots, p_T \in \mathcal{P}(A)$ and define new probability distributions on A^n by

$$p'_t(x^n) := \begin{cases} \frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}_{p_t,\delta}^n)} & \text{if } x^n \in \mathcal{T}_{p_t,\delta}^n, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

Define then for $z^n \in C^n$, $x^n \in A^n$

$$\tilde{Q}_{t,x^n}(z^n) = V_t^n(z^n|x^n) \cdot \mathbf{1}_{\mathcal{T}_{V_t,\delta}^n(x^n)}(z^n)$$

on C^n . Additionally, we set for $z^n \in C^n$

$$\Theta'_t(z^n) = \sum_{x^n \in \mathcal{T}_{p_t,\delta}^n} p'_t(x^n) \tilde{Q}_{t,x^n}(z^n). \quad (11)$$

Now let $S := \{z^n \in C^n : \Theta'_t(z^n) \geq \epsilon \alpha_t\}$ where $\epsilon = 2^{-nc'\delta^2}$ (cf. Lemma 3.1) and α_t is from (8) in Lemma 3.2 computed with respect to p_t and V_t . By lemma 3.2 the support of Θ'_t has cardinality $\leq \alpha_t^{-1}$ since for each $x^n \in \mathcal{T}_{p_t,\delta}^n$ it holds that $\mathcal{T}_{V_t,\delta}^n(x^n) \subset \mathcal{T}_{p_t V_t, 2|A|\delta}^n$, which implies that $\sum_{z^n \in S} \Theta_t(z^n) \geq 1 - 2\epsilon$, if

$$\begin{aligned} \Theta_t(z^n) &= \Theta'_t(z^n) \cdot \mathbf{1}_S(z^n) \quad \text{and} \\ Q_{t,x^n}(z^n) &= \tilde{Q}_{t,x^n}(z^n) \cdot \mathbf{1}_S(z^n). \end{aligned} \quad (12)$$

Now for each $t \in \theta$ define $J_n, L_{n,t}$ i.i.d. random variables $X_{j,l}^{(t)}$ with $j \in [J_n] := \{1, \dots, J_n\}$ and $l \in [L_{n,t}] := \{1, \dots, L_{n,t}\}$ each of them distributed according to p'_t with

$$J_n = \left\lfloor 2^{n[\min_{t \in \theta} (I(p_t, W_t) - I(p_t, V_t)) - \tau]} \right\rfloor \quad (13)$$

$$L_{n,t} = \left\lfloor 2^{n[I(p_t, V_t) + \frac{\tau}{4}]} \right\rfloor \quad (14)$$

for $\tau > 0$. Moreover we suppose that the random matrices $\{X_{j,l}^{(t)}\}_{j \in [J_n], l \in [L_{n,t}]}$ and $\{X_{j,l}^{(t')}\}_{j \in [J_n], l \in [L_{n,t}]}$ are independent for $t \neq t'$. Now it is obvious from (11) and the definition of the set S that for any $z^n \in S$ $\Theta_t(z^n) = \mathbb{E} Q_{t, X_{j,l}^{(t)}}(z^n) \geq \epsilon \alpha_t$ if \mathbb{E} is the expectation value with respect to the distribution p'_t . For the random variables $\beta_t^{-1} Q_{t, X_{j,l}^{(t)}}(z^n)$ define the event

$$\iota_j(t) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{t, X_{j,l}^{(t)}}(z^n) \in [(1 \pm \epsilon)\Theta_t(z^n)] \right\}, \quad (15)$$

and keeping in mind that $\Theta_t(z^n) \geq \epsilon \alpha_t$ for all $z^n \in S$ it follows that for all $j \in [J_n]$ and for all $t \in \theta$

$$\Pr\{(\iota_j(t))^c\} \leq 2|C|^n \exp \left(-L_{n,t} \frac{2^{-n[I(p_t, V_t) + g(\delta)]}}{3} \right) \quad (16)$$

by Lemma 3.4, Lemma 3.2, and our choice $\epsilon = 2^{-nc'\delta^2}$ with $g(\delta) := f_1(\delta) + f_2(\delta) + 3c'\delta^2$. Making $\delta > 0$ sufficiently small we have for all sufficiently large $n \in \mathbb{N}$

$$L_{n,t} 2^{-n[I(p_t, V_t) + g(\delta)]} \geq 2^{n\frac{\tau}{8}}.$$

Thus, for this choice of δ the RHS of (16) is double exponential in n uniformly in $t \in \theta$ and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large $n \in \mathbb{N}$. I.e.

$$\Pr\{(\iota_j(t))^c\} \leq \epsilon J_n^{-1} \quad \forall t \in \theta. \quad (17)$$

Let us turn now to the coding part of the problem. Let $p'_t \in \mathcal{P}(A^n)$ be given as in (10). We abbreviate $\mathcal{X} := \{X^{(t)}\}_{t \in \theta}$ for

the family of random matrices $X^{(t)} = \{X_{jl}^{(t)}\}_{j \in [J_n], l \in [L_{n,t}]}$ whose components are i.i.d. according to p_t . We will show now how the reliable transmission of the message $j \in [J_n]$ can be achieved when randomising over the index $l \in [L_{n,t}]$ without any attempt to decode the randomisation parameter at the legitimate receiver (see section IV-A). To this end let us define for each $j \in [J_n]$ a random set

$$D'_j(\mathcal{X}) := \bigcup_{s \in \theta} \bigcup_{k \in [L_{n,s}]} \mathcal{T}_{W_{s,\delta}}^n(X_{jk}^{(s)}),$$

and the subordinate random decoder $\{D_j(\mathcal{X})\}_{j \in [J_n]} \subseteq B^n$ is given by

$$D_j(\mathcal{X}) := D'_j(\mathcal{X}) \cap \left(\bigcup_{\substack{j' \in [J_n] \\ j' \neq j}} D'_{j'}(\mathcal{X}) \right)^c. \quad (18)$$

Consequently we can define the random average probabilities of error for a specific channel $t \in \theta$ by

$$\lambda_n^{(t)}(\mathcal{X}) := \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)}). \quad (19)$$

Now (18) implies for each $t \in \theta$ and $l \in [L_{n,t}]$

$$\begin{aligned} & W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)}) \\ & \leq W_t^{\otimes n} \left(\bigcap_{s \in \theta} \bigcap_{k \in [L_{n,s}]} (\mathcal{T}_{W_{s,\delta}}^n(X_{jk}^{(s)}))^c | X_{jl}^{(t)} \right) \\ & + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{s \in \theta} \sum_{k \in [L_{n,s}]} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | X_{jl}^{(t)}) \\ & \leq W_t^{\otimes n}((\mathcal{T}_{W_{t,\delta}}^n(X_{jl}^{(t)}))^c | X_{jl}^{(t)}) \\ & + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{s \in \theta} \sum_{k \in [L_{n,s}]} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | X_{jl}^{(t)}), \end{aligned} \quad (20)$$

where the second inequality follows by the monotonicity of the probability. By Lemma 3.1 and the independence of all involved random variables we obtain

$$\begin{aligned} & \mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)})) \\ & \leq (n+1)^{|A||B|} \cdot 2^{-nc\delta^2} \\ & + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{s \in \theta} \sum_{k \in [L_{n,s}]} \mathbb{E}_{X_{j'k}^{(s)}} \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | X_{jl}^{(t)}). \end{aligned}$$

(21)

We shall find now for $j' \neq j$ an upper bound on

$$\begin{aligned} & \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | X_{jl}^{(t)}) \\ & = \sum_{x^n \in A^n} p_t'(x^n) W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | x^n) \\ & \leq \sum_{x^n \in A^n} \frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}_{p_t,\delta}^n)} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | x^n) \\ & = \frac{q_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}))}{p_t^{\otimes n}(\mathcal{T}_{p_t,\delta}^n)}. \end{aligned} \quad (22)$$

By Lemma 3.1 and by Lemma 3.3 for any $t, s \in \theta$ we have

$$\begin{aligned} p_t^{\otimes n}(\mathcal{T}_{p_t,\delta}^n) & \geq 1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2} \\ q_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)})) & \leq (n+1)^{|A||B|} \cdot 2^{-n(I(p_s, W_s) - f(\delta))} \end{aligned} \quad (23)$$

with a universal $f(\delta) > 0$ satisfying $\lim_{\delta \rightarrow 0} f(\delta) = 0$ since $X_{j'k}^{(s)} \in \mathcal{T}_{p_s,\delta}^n$ with probability 1. Thus inserting this into (22) we obtain

$$\begin{aligned} & \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_{s,\delta}}^n(X_{j'k}^{(s)}) | X_{jl}^{(t)}) \\ & \leq \frac{(n+1)^{|A||B|}}{1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p_s, W_s) - f(\delta))} \end{aligned}$$

for all $s, t \in \theta$, all $j' \neq j$, and all $l \in [L_{n,t}], k \in [L_{n,s}]$. Now by defining $\nu_n(\delta) := (n+1)^{|A||B|} \cdot 2^{-nc\delta^2}$ and $\mu_n(\delta) := 1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}$ thus for each $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$ (21) and (22) lead to

$$\begin{aligned} & \mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)})) \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{s \in \theta} L_{n,s} 2^{-n(I(p_s, W_s) - f(\delta))} \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{s \in \theta} 2^{-n(I(p_s, W_s) - I(p_s, V_s) - f(\delta) - \frac{\tau}{4})} \\ & \leq \nu_n(\delta) \\ & + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot J_n \cdot 2^{-n(\min_{s \in \theta} (I(p_s, W_s) - I(p_s, V_s)) - f(\delta) - \frac{\tau}{4})} \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot 2^{-n(\tau - f(\delta) - \frac{\tau}{4})} \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot 2^{-n\frac{\tau}{2}} \end{aligned} \quad (24)$$

where we have used (14), (13), and we have chosen $\delta > 0$ small enough to ensure that $\tau - f(\delta) - \frac{\tau}{4} \geq \frac{\tau}{2}$. Defining $a = a(\delta, \tau) := \frac{\min\{c\delta^2, \frac{\tau}{4}\}}{2}$ we can find $n(\delta, \tau, |A|, |B|) \in \mathbb{N}$ such that for all $n \geq n(\delta, \tau, |A|, |B|)$

$$\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)})) \leq T \cdot 2^{-na}$$

holds for all $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$. Consequently, for any $t \in \theta$ we obtain

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(t)}(\mathcal{X})) \leq T \cdot 2^{-na}.$$

Additionally we define for any $t \in \theta$ an event

$$\iota_0(t) = \{\lambda_n^{(t)}(\mathcal{X}) \leq \sqrt{T} 2^{-n\frac{a}{2}}\}. \quad (25)$$

Then using the Markov inequality applied to $\lambda_n^{(t)}(\mathcal{X})$ along with (25), we obtain that

$$\Pr\{(\iota_0(t))^c\} \leq \sqrt{T} 2^{-n\frac{a}{2}}. \quad (26)$$

Set

$$\iota := \bigcap_{t \in \theta} \bigcap_{k=0}^{J_n} \iota_k(t) \quad (27)$$

Then with (17), (26), and applying the union bound we obtain

$$\begin{aligned} \Pr\{\iota^c\} & \leq \sum_{t \in \theta} \sum_{k=0}^{J_n} \Pr\{(\iota_k(t))^c\} \leq T \cdot \epsilon + T^{\frac{3}{2}} \cdot 2^{-n\frac{a}{2}} \\ & \leq T^2 \cdot 2^{-nc''} \end{aligned}$$

for a suitable positive constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$.

Hence, we have shown that for each $t \in \theta$ there exist realisations $\{(x_{jl}^{(t)})_{j \in [J_n], l \in [L_{n,t}]} : t \in \theta\} \in \iota$ of \mathcal{X} . Now, denoting by $\|\cdot\|$ the variational distance

$$\|p - q\| := \sum_{x \in A} |p(x) - q(x)|$$

for $p, q \in A$, we show that the secrecy level is fulfilled uniformly in $t \in \theta$ for any particular $\{(x_{jl}^{(t)})_{j \in [J_n], l \in [L_{n,t}]} : t \in \theta\} \in \iota$.

$$\begin{aligned} & \left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} V_t^n(\cdot | x_{jl}^{(t)}) - \Theta_t(\cdot) \right\| \\ & \leq \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \|V_t^n(\cdot | x_{jl}^{(t)}) - \tilde{Q}_{t,x_{jl}^{(t)}}(\cdot)\| \\ & + \left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} (\tilde{Q}_{t,x_{jl}^{(t)}}(\cdot) - Q_{t,x_{jl}^{(t)}}(\cdot)) \right\| \\ & + \left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{t,x_{jl}^{(t)}}(\cdot) - \Theta_t(\cdot) \right\| \leq 5\epsilon. \end{aligned} \quad (28)$$

In the first term the functions $V_t^n(\cdot | x_{jl}^{(t)})$ and $\tilde{Q}_{t,x_{jl}^{(t)}}(\cdot)$ differ if $z^n \notin \mathcal{T}_{p_t V_t, 2|A|\delta}^n$, so it makes a contribution of ϵ to the bound. In the second term \tilde{Q}_t and Q_t are different for $z^n \notin S$ and because $\iota_j(t)$ and $\sum_{z^n \in S} \Theta_t(z^n) \geq 1 - 2\epsilon$ imply that

$$\frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \sum_{z^n \in S} Q_{t,x_{jl}^{(t)}}(z^n) \geq 1 - 3\epsilon,$$

the second term is bounded by 3ϵ . The third term is bounded by ϵ which follows directly from (15).

For any $\{(x_{jl}^{(t)})_{j \in [J_n], l \in [L_{n,t}]} : t \in \theta\} \in \iota$ with the corresponding decoding sets $\{D_j : j \in [J_n]\}$ it follows by construction that

$$\frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \leq \sqrt{T} \cdot 2^{-na'} \quad (29)$$

is fulfilled for all $t \in \theta$ with $a' > 0$, which means that we have found a (n, J_n) code with average error probability tending to zero for $n \in \mathbb{N}$ sufficiently large for any channel realisation. Now by a standard expurgation scheme we show that this still holds for the maximum error probability. We define the set

$$G_t := \{j \in J_n : \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \leq \sqrt{\eta}\} \quad (30)$$

with $\eta := \sqrt{T} \cdot 2^{-na'}$ and denote its complement as $B_t := G_t^c$ and the union of all complements as $B = \bigcup_{t \in \theta} B_t$. Then (29) and (30) imply that

$$\eta \geq \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \geq \frac{|B_t|}{J_n} \sqrt{\eta}$$

for all $t \in \theta$ and by the union bound it follows that

$$|B| \leq \sum_{t \in \theta} |B_t| \leq T \cdot \sqrt{\eta} \cdot J_n.$$

After removing all $j \in B$ (which are at most a fraction of $T^{\frac{5}{4}} 2^{-n \frac{a'}{2}}$ of J_n) and relabeling we obtain a new (n, \tilde{J}_n) code $(E_j, D_j)_{j \in [\tilde{J}_n]}$ without changing the rate. The maximum error probability of the new code fulfills for sufficiently large $n \in \mathbb{N}$

$$\max_{t \in \theta} \max_{j \in [\tilde{J}_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \leq T^{\frac{1}{4}} \cdot 2^{-n \frac{a'}{2}}.$$

On the other hand, if we set

$$\hat{V}_t^n(z^n | (j, l)) := V_t^n(z^n | x_{jl}^{(t)}) \quad (31)$$

and further define

$$\hat{V}_{t,j}^n(z^n) = \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \hat{V}_t^n(z^n | (j, l)), \quad (32)$$

$$\bar{V}_t^n(z^n) = \frac{1}{\tilde{J}_n} \sum_{j=1}^{\tilde{J}_n} \hat{V}_{t,j}^n(z^n), \quad (33)$$

we obtain that

$$\begin{aligned} \|\hat{V}_{t,j}^n - \bar{V}_t^n\| & \leq \|\hat{V}_{t,j}^n - \Theta_t\| + \|\Theta_t - \bar{V}_t^n\| \\ & \leq 10\epsilon, \end{aligned}$$

for all $j \in [\tilde{J}_n], t \in \theta$ with $\epsilon = 2^{-nc'\delta^2}$ where we have used the convexity of the variational distance and (28) which still applies by our expurgation procedure. For a uniformly distributed random variable J taking values in the set $\{1, \dots, \tilde{J}_n\}$ we obtain with Lemma 2.7 of [11] (uniform continuity of the entropy function)

$$\begin{aligned} I(J; Z_t^n) & = \sum_{j=1}^{\tilde{J}_n} \frac{1}{\tilde{J}_n} (H(\bar{V}_t^n) - H(\hat{V}_{t,j}^n)) \\ & = H(Z_t^n) - H(Z_t^n | J) \\ & \leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log |C| \end{aligned}$$

uniformly in $t \in \theta$ (for $10\epsilon \leq e^{-1}$). Hence the strong secrecy level of the definition 2.1 holds uniformly in $t \in \theta$. Using standard arguments (cf. [11] page 409) we then have shown the achievability of the secrecy rate

$$R_S = \min_{t \in \theta} \max_{V \rightarrow X \rightarrow (YZ)_t} (I(V, Y_t) - I(V, Z_t)). \quad (34)$$

Remark. Note that in the case that $\mathfrak{W} := \{W_t, V_s : t = 1, \dots, T, s = 1, \dots, S\}$ with $S \neq T$ and the pair (s, t) known to the transmitter prior to transmission nothing new happens. A slight modification of the arguments presented above shows that

$$C_{S,CSI}(\mathfrak{W}) = \min_{(t,s)} \max_{V \rightarrow X \rightarrow (Y_t Z_s)} (I(V, Y_t) - I(V, Z_s)).$$

C. No CSI

In the previous section we have assumed that the channel state is known to the transmitter. We now consider the case where neither the transmitter nor the receiver has knowledge of the channel state. We will prove that

Theorem 3.6: For the secrecy capacity $C_S(\mathfrak{W})$ of the compound wiretap channel \mathfrak{W} without CSI it holds that

$$C_S(\mathfrak{W}) \geq \max_{p \in \mathcal{P}(A)} (\min_{t \in \theta} I(p, W_t) - \max_{t \in \theta} I(p, V_t)).$$

Proof: Caused by the lack of channel knowledge we use a stochastic encoder independent of the channel realisation. For any $p \in \mathcal{P}(A)$ let $p' \in \mathcal{P}(A^n)$ be the distribution given by

$$p'(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}_{p,\delta}^n)} & \text{if } x^n \in \mathcal{T}_{p,\delta}^n, \\ 0 & \text{otherwise.} \end{cases}$$

Then analogously to the case with CSI we define $\tilde{Q}_{t,x^n}(z^n)$, $Q_{t,x^n}(z^n)$, and $\Theta'_t(z^n)$, $\Theta_t(z^n)$ for $z^n \in C^n$ but now with respect to the distribution p' . Consequently, $\Theta'(\cdot)$ has support only on $\mathcal{T}_{pV_t, 2|A|\delta}^n$, and $Q_{t,x^n}(\cdot)$ and $\Theta(\cdot)$ only on the set S . Furthermore $\Theta(z^n) \geq \epsilon \alpha_t$ for all $z^n \in S$. Now define $J_n \cdot L_n$ i.i.d random variables X_{jl} according to the distribution p' independent of $t \in \theta$ with $j \in [J_n]$ and $l \in [L_n]$ with

$$J_n = \lfloor 2^{n[\min_t I(p, W_t) - \max_t I(p, V_t) - \tau]} \rfloor \quad (35)$$

$$L_n = \lfloor 2^{n[\max_t I(p, V_t) + \frac{\tau}{4}]} \rfloor \quad (36)$$

for $\tau > 0$. Now because $\Theta_t(z^n) = \mathbb{E} Q_{t,X_{jl}} \geq \epsilon \alpha_t$ for all $z^n \in S$ we define the event $\iota_j(t)$ as in (15) for the random variables $\beta_t^{-1} Q_{t,X_{jl}}$

$$\iota_j(t) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{t,X_{jl}}(z^n) \in [(1 \pm \epsilon) \Theta_t(z^n)] \right\},$$

but considering the difference that the random variables X_{jl} are independent of the channel state. Then analogously to (16) we obtain that

$$\Pr\{(\iota_j(t))^c\} \leq 2|C|^n \exp\left(-L_n \frac{2^{-n(I(p, V_t) + g(\delta))}}{3}\right)$$

by Lemma 3.4 and Lemma 3.2. Notice that, because the sender does not know which channel is used, we need the maximum in the definition of L_n . Thus the right-hand side is a double exponential in n and can be made smaller than ϵJ_n^{-1} for all j and for all $t \in \theta$ and sufficiently large n .

Now let J_n and L_n be defined as stated above, and let $X^n = \{X_{jl}\}_{j \in [J_n], l \in [L_n]}$ be the set of i.i.d. random variables each of them distributed according to p' independent of $t \in \theta$. As in the case of CSI we can show that reliable transmission of the message $j \in [J_n]$ can be achieved. To this end define now the random decoder $\{D_j(X^n)\}_{j \in [J_n]} \subseteq B^n$ as in (18) but with

$$D'_j(X^n) := \bigcup_{s \in \theta} \bigcup_{k \in [L_n]} \mathcal{T}_{W_s, \delta}^n(X_{jk}),$$

and the random average probabilities of error for a specific channel $\lambda_n^{(t)}(X^n)$ as in (19). Notice that now both X^n and L_n do not depend on $t \in \theta$ and this holds throughout the entire proof. Then we can give the bound in (20) now by

$$\begin{aligned} & W_t^{\otimes n}((D_j(X^n))^c | X_{jl}) \\ & \leq W_t^{\otimes n}((\mathcal{T}_{W_t, \delta}^n(X_{jl}))^c | X_{jl}) \\ & + \sum_{j' \in [J_n]} \sum_{s \in \theta} \sum_{k \in [L_n]} W_t^{\otimes n}(\mathcal{T}_{W_s, \delta}^n(X_{j'k}) | X_{jl}) \end{aligned}$$

We can bound the first term in the inequality by $\nu_n(\delta) := (n+1)^{|A||B|} \cdot 2^{-nc\delta^2}$ (see (21)). If we average over all codebooks we get

$$\begin{aligned} & \mathbb{E}_{X^n} (W_t^{\otimes n}((D_j(X^n))^c | X_{jl})) \leq \\ & \nu_n(\delta) + \sum_{j' \in [J_n]} \sum_{s \in \theta} \sum_{k \in [L_n]} \mathbb{E}_{X_{j'k}} \mathbb{E}_{X_{jl}} W_t^{\otimes n}(\theta_{W_s, \delta}^n(X_{j'k}) | X_{jl}). \end{aligned}$$

By the same reasoning as in (22) and (23) we can give an upper bound on

$$\begin{aligned} & \mathbb{E}_{X_{jl}} W_t^{\otimes n}(\mathcal{T}_{W_s, \delta}^n(X_{j'k}) | X_{jl}) \\ & \leq \frac{q_t^{\otimes n}(\mathcal{T}_{W_s, \delta}^n(X_{j'k}))}{p^{\otimes n}(\mathcal{T}_{p, \delta}^n)} \\ & \leq \frac{(n+1)^{|A||B|}}{1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p, W_s) - f(\delta))} \end{aligned}$$

for all $t \in \theta$, all $j' \neq j$ and all $k, l \in [L_n]$ with a universal $f(\delta) > 0$ satisfying $\lim_{\delta \rightarrow 0} f(\delta) = 0$. $q_t^{\otimes n}$ denotes the output distribution generated by the conditional distribution $W_t^{\otimes n}$ and the input distribution $p^{\otimes n}$. Additionally we define $\mu_n(\delta) := 1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}$. Then (24) changes to

$$\begin{aligned} & \mathbb{E}_{X^n} (W_t^{\otimes n}((D_j(X^n))^c | X_{jl})) \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot J_n L_n \cdot 2^{-n(\min_s I(p, W_s) - f(\delta))} \\ & \leq \nu_n(\delta) + T \cdot 2^{-n\frac{\tau}{2}} \end{aligned}$$

by the definition of J_n and L_n and by choosing $\delta > 0$ small enough that $\tau - \frac{\tau}{4} - f(\delta) \geq \frac{\tau}{2}$. Now by defining $a := \frac{\min\{c\delta^2, \frac{\tau}{2}\}}{2}$ and the definition of the error probability the last inequality results in the upper bound

$$\mathbb{E}_{X^n} (\lambda_n^{(t)}(X^n)) \leq T \cdot 2^{-na}$$

for any $t \in \theta$ and $n \in \mathbb{N}$ large enough.

Now we define the event $\iota_0(t)$ for any $t \in \theta$ and the event ι as in (25) and (27) but with the difference that the input is independent of the channel realisation. So by the same reasoning we end in

$$\Pr\{\iota^c\} \leq T^2 \cdot 2^{-nc''}$$

for a constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$, which implies that there exist realisations $\{x_{jl}\}$ of $\{X_{jl}\}$ such that $x_{jl} \in \iota$ for all $j \in [J_n]$ and $l \in [L_n]$. Then analogously to (28) we get for any channel $t \in \theta$

$$\left\| \frac{1}{L_n} \sum_{l=1}^{L_n} V_t^n(\cdot | x_{jl}) - \Theta_t(\cdot) \right\| \leq 5\epsilon$$

differs from the former only by L_n in place of $L_{n,t}$. Hence, following the same arguments subsequent to (29), we have shown that there is a sequence of (n, \tilde{J}_n) codes for which

$$\max_{t \in \theta} \max_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} W_t^{\otimes n}(D_j^c | x_{jl}) \leq T^{\frac{1}{4}} \cdot 2^{-n\frac{a'}{2}}$$

holds for sufficiently large $n \in \mathbb{N}$, and the strong secrecy level is fulfilled for every channel $t \in \theta$ by

$$\|\hat{V}_{t,j}^n - \bar{V}_t^n\| \leq 6\epsilon$$

($\hat{V}_{t,j}^n, \bar{V}_t^n$ defined as in (32), (33)) and thus by

$$I(J; Z_t^n) \leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log |C|$$

which tends to zero for $n \rightarrow \infty$ uniformly in $t \in \theta$. ■

We turn now to the converse of Theorem 3.6. Actually, we give only a multiletter formula of the upper bound of the secrecy rates. First we need the following lemma.

Lemma 3.7: Let $\mathfrak{W} = \{(W_t, V_t) : t \in \theta\}$ be an arbitrary compound wiretap channel without CSI. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n))$$

exists and we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n)) \\ &= \sup_{n \in \mathbb{N}} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n)). \end{aligned}$$

Proof: The proof is based on Fekete's lemma [13]. Consequently, if we apply the lemma to the sequence $(a_n)_{n \in \mathbb{N}}$ defined by

$$a_n := \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n))$$

it suffices to show that the inequality

$$a_{n+m} \geq a_n + a_m$$

holds for all $n, m \in \mathbb{N}$. This will be done by considering two independent Markov chains $U_1 \rightarrow X^n \rightarrow (Y_t^n, Z_t^n)$ and $U_2 \rightarrow \hat{X}^m \rightarrow (\hat{Y}_t^m, \hat{Z}_t^m)$ and setting $U := (U_1, U_2)$, $X^{n+m} := (X^n, \hat{X}^m)$, and $(Y_t^{n+m}, Z_t^{n+m}) := ((Y_t^n, \hat{Y}_t^m), (Z_t^n, \hat{Z}_t^m))$. Then by the definition of a_n

$$\begin{aligned} a_{n+m} &\geq \inf_{t \in \theta} I(U; Y_t^{n+m}) - \sup_{t \in \theta} I(U; Z_t^{n+m}) \\ &\geq \inf_{t \in \theta} I(U_1; Y_t^n) + \inf_{t \in \theta} I(U_2; \hat{Y}_t^m) \\ &\quad - \sup_{t \in \theta} I(U_1; Z_t^n) - \sup_{t \in \theta} I(U_2; \hat{Z}_t^m). \end{aligned}$$

By the independence of the two Markov chains mentioned above and because apart from that these Markov chains were arbitrary we can conclude that

$$a_{n+m} \geq a_n + a_m$$

holds for all $n, m \in \mathbb{N}$. ■

Proposition 3.8: The secrecy capacity of the compound wiretap channel in the case of no CSI $C_S(\mathfrak{W})$ is upper bounded by

$$C_S(\mathfrak{W}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n)).$$

Proof: Let $(\mathcal{C}_n)_{n \in \mathbb{N}}$ be any sequence of (n, J_n) codes such that with

$$\sup_{t \in \theta} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n | j) W_t^{\otimes n} (D_j^c | x^n) =: \varepsilon_{1,n}, \quad (37)$$

and

$$\sup_{t \in \theta} I(J; Z_t^n) =: \varepsilon_{2,n}$$

it holds that $\lim_{n \rightarrow \infty} \varepsilon_{1,n} = 0$ and $\lim_{n \rightarrow \infty} \varepsilon_{2,n} = 0$, where J denotes the random variable which is uniformly distributed on the message set $\{1, \dots, J_n\}$. Let us denote by \hat{J} the random variable with values in $\{1, \dots, J_n\}$ determined by the Markov chain $J \rightarrow X^n \rightarrow Y_t^n \rightarrow \hat{J}$ where the first transition is governed by E , the second by $W_t^{\otimes n}$, and the last by the decoding rule. Then we have for any $t \in \theta$

$$\begin{aligned} \log J_n &= H(J) \\ &= I(J; \hat{J}) + H(J | \hat{J}) \\ &\leq I(J; Y_t^n) + H(J | \hat{J}), \end{aligned} \quad (38)$$

where the inequality follows from the data processing inequality. Then using Fano's inequality we find that

$$H(J | \hat{J}) \leq 1 + \varepsilon_{1,n} \log J_n$$

with (37). Thus we can rewrite inequality (38) as

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J; Y_t^n) + 1$$

for all $t \in \theta$. On the other hand we have for every $t \in \theta$

$$I(J; Y_t^n) = I(J; Y_t^n) - \sup_{t \in \theta} I(J; Z_t^n) + \varepsilon_{2,n}$$

where we have used the validity of the secrecy criterion stated above. Then the last two inequalities imply that for any $t \in \theta$

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J; Y_t^n) - \sup_{t \in \theta} I(J; Z_t^n) + \varepsilon_{2,n}. \quad (39)$$

Since the LHS of (39) does not depend on t we arrive at

$$\begin{aligned} & (1 - \varepsilon_{1,n}) \log J_n \\ & \leq \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta} I(U; Z_t^n)) + \varepsilon_{2,n}, \end{aligned}$$

which concludes the proof after dividing by $n \in \mathbb{N}$, taking lim sup and taking into account the assertion of Lemma 3.7. ■

Remark. Following the same arguments subsequent to (34) concerning the use of the channels defined by $P_{Y_t|T} = W_t \cdot P_{X|T}$ and $P_{Z_t|T} = V_t \cdot P_{X|T}$ instead of W_t and V_t and applying the assertion of Theorem 3.6 to the n -fold product of channels W_t and V_t , we can give the coding theorem for the multiletter case. The capacity of the compound wiretap channel in the case of no CSI is

$$C_S(\mathfrak{W}) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_t^n} (\inf_{t \in \theta} I(U; Y_t^n) - \sup_{t \in \theta_y} I(U; Z_t^n)).$$

Let us consider now the case $\mathfrak{W} := \{W_t, V_s : t = 1, \dots, T, s = 1, \dots, S\}$ with $S \neq T$ and the pair (s, t) unknown to both the transmitter and the legitimate receiver. Additionally we assume that each V_s is a degraded version of every W_t , which is characterised by

$$V_s(z|x) = \sum_{y \in B} W_t(y|x) D_{(t,s)}(z|y), \quad (40)$$

for all $x \in A, z \in C$, if D is defined as the stochastic matrix $D : B \rightarrow \mathcal{P}(C)$. Then we have the following

Lemma 3.9: Let $p \in \mathcal{P}(A)$, $W : A \rightarrow \mathcal{P}(B)$, $V : A \rightarrow \mathcal{P}(C)$, and assume that V is a degraded version of W . Then $I(X; Y|Z)$ is a concave with respect to the input distribution $p_X = p$.

Proof: Let X, Y, Z be random variables with values in A, B, C respectively distributed according to

$$\Pr(X = x, Y = y, Z = z) := p_{XYZ}(x, y, z) = p(x)W(y|x)D(z|y) \quad (41)$$

for all $x \in A, y \in B, z \in C$. Because

$$I(X; Y|Z) = H(Y|Z) - H(Y|X, Z)$$

the proof is based on the two assertions

- 1) $H(Y|Z)$ depends concavely on p_X , and
- 2) $H(Y|X, Z)$ is an affine function of p_X .

First, $H(Y|Z)$ is a concave function with respect to p_{YZ} by the log-sum inequality (cf. [11] Lemma 3.1). Then because p_{XYZ} depends affinely on p_X by (41), so does p_{YZ} , and the first assertion follows. For the second consider that (40) and (41) imply that

$$p_{Y|X, Z}(y|x, z) = \frac{W(y|x)D(z|y)}{V(z|x)}$$

for every input distribution p_X , any $y \in B$ and all $x \in A, z \in C$ with $p_{XZ}(x, z) > 0$. Then we have

$$H(Y|X, Z) = \sum_{x \in A, z \in C} p_{XZ}(x, z) H\left(\frac{W(\cdot|x)D(\cdot|z)}{V(z|x)}\right)$$

showing that $H(Y|X, Z)$ is an affine function of p_{XZ} which in turn depends affinely on p_X . ■

Now because the random variables X, Y_t, Z_s (Y_t, Z_s the channel outputs of W_t and V_s resp.) form a Markov chain for all $t \in \theta$ and $s \in \mathcal{S}$, we obtain that

$$I(X; Y_t|Z_s) = I(X, Y_t) - I(X, Z_s). \quad (42)$$

By virtue of Theorem 2 of [14] we can show that for the secrecy rate it holds that

$$R_S \leq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_{i,t}|Z_{i,s}) + \epsilon'$$

for any channel $(t, s) \in \theta \times \mathcal{S}$ and $\epsilon' > 0$. The concavity of $I(X; Y_t|Z_s)$ with respect to the input distributions $p \in \mathcal{P}(A)$ together with (42) then imply the converse part of Theorem 3.6, that

$$R_S \leq \max_{p \in \mathcal{P}(A)} \min_{(t,s)} (I(p, W_t) - I(p, V_s)).$$

Now we can state the following

Proposition 3.10: If V_s is a degraded version of W_t for all $s \in \mathcal{S}$ and $t \in \theta$ the capacity of the compound wiretap channel is given by

$$\begin{aligned} C_S(\mathfrak{W}) &= \max_{p \in \mathcal{P}(A)} \min_{(t,s)} (I(p, W_t) - I(p, V_s)) \\ &= \max_{p \in \mathcal{P}(A)} (\min_t I(p, W_t) - \max_s I(p, V_s)). \end{aligned}$$

Remark. This result was obtained in [1] with a weaker notion of secrecy.

D. Channel state to the legitimate receiver is known at the transmitter (CSI_t)

We now consider the case, in which the transmitter has knowledge of the channel state to the legitimate receiver $t \in \theta$ but the channel state to the eavesdropper $s \in \mathcal{S}$ is unknown. Consequently we get for each $t \in \theta$ possible channel realisations $\mathfrak{W}_t := \{(W_t, V_s) : s = 1, \dots, S\}$. Then we can describe the compound channel as $\mathfrak{W} = \cup_{t \in \theta} \mathfrak{W}_t$.

Theorem 3.11: For the secrecy capacity $C_{S, CSI_t}(\mathfrak{W})$ of the compound wiretap channel with CSI to the legitimate receiver it holds that

$$C_{S, CSI_t}(\mathfrak{W}) \geq \min_{t \in \theta} \max_{p \in \mathcal{P}(A)} (I(p, W_t) - \max_{s \in \mathcal{S}} I(p, V_s)).$$

Proof: Adapted to the channel realisation W_t define

$$p'_t(x^n) := \begin{cases} \frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}_{p_t, \delta}^n)} & \text{if } x^n \in \mathcal{T}_{p_t, \delta}^n, \\ 0 & \text{otherwise.} \end{cases} \quad (43)$$

for arbitrary input distributions $p_1, \dots, p_T \in \mathcal{P}(A)$. Now define for $z^n \in C^n$ and $s \in \mathcal{S}$

$$\tilde{Q}_{s, x^n}(z^n) = V_s^n(z^n | x^n) \cdot \mathbf{1}_{\mathcal{T}_{V_s, \delta}^n(x^n)}(z^n)$$

on C^n . Additionally, we set for $z^n \in C^n$

$$\Theta'_s(z^n) = \sum_{x^n \in \mathcal{T}_{p_t, \delta}^n} p'_t(x^n) \tilde{Q}_{s, x^n}(z^n).$$

Now let $S := \{z^n \in C^n : \Theta'_s(z^n) \geq \epsilon \alpha_{t,s}\}$ where $\epsilon = 2^{-nc'\delta^2}$ and $\alpha_{t,s}$ is from (8) similar to the former cases but computed with respect to p_t and V_s . Then the support of Θ'_s has cardinality $\leq \alpha_{t,s}^{-1}$, which implies that $\sum_{z^n \in S} \Theta'_s(z^n) \geq 1 - 2\epsilon$. Analogously to (12) define $\Theta_s(z^n)$ and $Q_{s, x^n}(z^n)$ with support on S and further

$$J_n = \lfloor 2^{n[\min_t (I(p_t, W_t) - \max_s I(p_t, V_s)) - \tau]} \rfloor \quad (44)$$

$$L_{n,t} = \lfloor 2^{n[\max_s I(p_t, V_s) + \frac{\tau}{4}]} \rfloor. \quad (45)$$

As in the case of CSI define random matrices $\{X_{jl}^{(t)}\}_{j \in [J_n], l \in [L_{n,t}]}$ such that the random variables $X_{jl}^{(t)}$ where i.i.d. according to p'_t . We suppose additionally that $\{X_{jl}^{(t)}\}_{j,l}$ and $\{X_{jl}^{(t')}\}_{j,l}$ are independent for $t \neq t'$. For any $z^n \in S$ it follows that $\Theta_s(z^n) = \mathbb{E} Q_{s, X_{jl}^{(t)}}(z^n) \geq \epsilon \alpha_{t,s}$, if \mathbb{E} is the expectation value with respect to the distribution p'_t . For the random variables $\beta_{t,s}^{-1} Q_{s, X_{jl}^{(t)}}(z^n)$ define the event

$$\iota_j(s, t) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{s, X_{jl}^{(t)}}(z^n) \in [(1 \pm \epsilon) \Theta_s(z^n)] \right\}.$$

Then it follows that for all $j \in [J_n]$ and for all $s \in \mathcal{S}$ it holds for each $t \in \theta$

$$\Pr\{(\iota_j(s, t))^c\} \leq 2|C|^n \exp\left(-L_{n,t} \frac{2^{-n[I(p_t, V_s) + g(\delta)]}}{3}\right)$$

by Lemma 3.4, Lemma 3.2. Thus the RHS is double exponential in n uniformly in $s \in \mathcal{S}, t \in \theta$ (guaranteed by the maximum in s in the definition of $L_{n,t}$) and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large n . Now the coding part of the problem is similar to

the case with CSI. Let $p'_t \in \mathcal{P}(A^n)$ be given as in (43). We abbreviate $\mathcal{X} := \{X^{(t)}\}_{t \in \theta}$ for the family of random matrices $X^{(t)} = \{X_{jl}^{(t)}\}_{j \in [J_n], l \in [L_{n,t}]}$ whose components are i.i.d. according to p'_t . We will show how reliable transmission of the message $j \in [J_n]$ can be achieved. To this end define now the random decoder $\{D_j(\mathcal{X})\}_{j \in [J_n]} \subseteq B^n$ as in (18) and with

$$D'_j(\mathcal{X}) := \bigcup_{r \in \theta} \bigcup_{k \in [L_{n,r}]} \mathcal{T}_{W_r, \delta}^n(X_{jk}^{(r)}),$$

and the random average probabilities of error for a specific channel $\lambda_n^{(t)}(\mathcal{X})$ as in (19) by

$$\lambda_n^{(t)}(\mathcal{X}) := \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)}).$$

As in (20) we get for each $t \in \theta$ and $l \in [L_{n,t}]$

$$\begin{aligned} & W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)}) \\ & \leq W_t^{\otimes n}((\mathcal{T}_{W_t, \delta}^n(X_{jl}^{(t)}))^c | X_{jl}^{(t)}) \\ & + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{r \in \theta} \sum_{k \in [L_{n,r}]} W_t^{\otimes n}(\mathcal{T}_{W_r, \delta}^n(X_{j'k}^{(r)}) | X_{jl}^{(t)}), \end{aligned}$$

Then by Lemma 3.1 we can bound the first term of the right hand side, such that together with the independence of all involved random variables we end up with

$$\begin{aligned} & \mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)})) \\ & \leq (n+1)^{|A||B|} \cdot 2^{-nc\delta^2} \\ & + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{r \in \theta} \sum_{k \in [L_{n,r}]} \mathbb{E}_{X_{j'k}^{(r)}} \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_r, \delta}^n(X_{j'k}^{(r)}) | X_{jl}^{(t)}). \end{aligned} \quad (46)$$

We shall find now for $j' \neq j$ by the same reasoning as in (22) and (23) an upper bound on

$$\begin{aligned} & \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_r, \delta}^n(X_{j'k}^{(r)}) | X_{jl}^{(t)}) \\ & \leq \frac{q_t^{\otimes n}(\mathcal{T}_{W_r, \delta}^n(X_{j'k}^{(r)}))}{p_t^{\otimes n}(\mathcal{T}_{p_t, \delta}^n)} \\ & \leq \frac{(n+1)^{|A||B|}}{1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p_r, W_r) - f(\delta))} \end{aligned}$$

for all $r, t \in \theta$, all $j' \neq j$, and all $l \in [L_{n,t}], k \in [L_{n,r}]$. Now by defining $\nu_n(\delta) := (n+1)^{|A||B|} \cdot 2^{-nc\delta^2}$ and $\mu_n(\delta) := 1 - (n+1)^{|A|} \cdot 2^{-nc\delta^2}$ thus for each $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$ (46) and the last inequality leads to

$$\begin{aligned} & \mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)})) \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{r \in \theta} L_{n,r} 2^{-n(I(p_r, W_r) - f(\delta))} \\ & \leq \nu_n(\delta) \\ & + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T J_n \cdot 2^{-n(\min_{r \in \theta} (I(p_r, W_r) - \max_s I(p_r, V_s)) - f(\delta) - \frac{\tau}{4})} \\ & \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot 2^{-n\frac{\tau}{2}} \end{aligned}$$

where we have used the definitions of J_n and $L_{n,r}$ and we have chosen $\delta > 0$ small enough to ensure that $\tau - f(\delta) -$

$\frac{\tau}{4} \geq \frac{\tau}{2}$. Defining $a = a(\delta, \tau) := \frac{\min\{c\delta^2, \frac{\tau}{2}\}}{2}$ we can find $n(\delta, \tau, |A|, |B|) \in \mathbb{N}$ such that for all $n \geq n(\delta, \tau, |A|, |B|)$ we end in

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(t)}(\mathcal{X})) \leq T \cdot 2^{-na}.$$

for any $t \in \theta$. To give a bound on the average probability of error we define the event $\iota_0(t)$ for any $t \in \theta$ as in (25) and the event

$$\iota := \bigcap_{t \in \theta} \bigcap_{s \in \mathcal{S}} \bigcap_{k=0}^{J_n} \iota_k(t, s)$$

differs from (27) only by the intersection of the unknown channel states $s \in \mathcal{S}$. Thus we can conclude that

$$\begin{aligned} \Pr\{\iota^c\} & \leq S \cdot T \cdot \epsilon + S \cdot T^{\frac{3}{2}} \cdot 2^{-n\frac{a}{2}} \\ & \leq S \cdot T^2 \cdot 2^{-nc''} \end{aligned}$$

holds for a suitable positive constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$, and we have shown that for each $t \in \theta$ there exist realisations $\{(x_{jl}^{(t)})_{j \in [J_n], l \in [L_{n,t}]} : t \in \theta\} \in \iota$ of \mathcal{X} . By the same reasoning as in (28) we get for any channel realisation $t \in \theta$ to the legitimate receiver

$$\left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} V_s^n(\cdot | x_{jl}^{(t)}) - \Theta_s(\cdot) \right\| \leq 5\epsilon$$

for each of the unknown channels $s \in \mathcal{S}$ to the eavesdropper. Now, because for any $t \in \theta$ we have a different codeword set $\{x_{jl}^{(t)}\}$, we slightly change the definition in (31) to

$$\hat{V}_{(s,t)}^n(z^n | (j, l)) := V_s^n(z^n | x_{jl}^{(t)})$$

and accordingly to $\hat{V}_{(s,t),j}^n$ and $\bar{V}_{(s,t)}^n$ in (32), (33) in that way, that these distributions are defined separately for each codeword set $t \in \theta$. Thus we get, that

$$\|\hat{V}_{(s,t),j}^n - \bar{V}_{(s,t)}^n\| \leq 10\epsilon$$

is fulfilled for all $s \in \mathcal{S}$ for each individual channel $t \in \theta$ to the legitimate receiver.

Hence, using the same expurgation scheme as in the previous sections we have shown that there is a sequence of (n, \tilde{J}_n) codes for which

$$\max_{t \in \theta} \max_{j \in [\tilde{J}_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \leq T^{\frac{1}{4}} \cdot 2^{-n\frac{a}{2}}$$

holds for sufficiently large $n \in \mathbb{N}$, and the strong secrecy level is fulfilled for every channel $t \in \theta$ by

$$I(J; Z_s^n) \leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log |C|$$

which tends to zero for $n \rightarrow \infty$ for all channels $s \in \mathcal{S}$ to the eavesdropper. Thus we have shown that

$$R_S = \min_{t \in \theta} \max_{p \in \mathcal{P}(A)} (I(p, W_t) - \max_{s: (s,t) \in \mathcal{S} \times \theta} I(p, V_s))$$

is an achievable secrecy rate for the compound wiretap channel $\cup_{t \in \theta} \mathfrak{W}_t$ in the case where the channel state to the legitimate receiver is known at the transmitter. ■

Remark. By considering the converse of Theorem 3.11, we get for each $t \in \theta$ possible channel realisations $\mathfrak{W}_t := \{(W_t, V_s) : s = 1, \dots, S\}$. Then we can describe the compound channel as

$\mathfrak{W} = \cup_{t \in \theta} \mathfrak{W}_t$. In accordance to the case of no CSI for each $t \in \theta$ we obtain that

$$C_S(\mathfrak{W}_t) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{U \rightarrow X^n \rightarrow Y_t^n Z_s^n} (I(U; Y_t^n) - \sup_{s \in \mathcal{S}} I(U; Z_s^n)).$$

Proposition 3.12: The secrecy capacity of the compound wiretap channel in the case where only the channel state to the legitimate receiver is known at the transmitter $C_{S,CSI_t}(\mathfrak{W})$ is given by

$$C_{S,CSI_t}(\mathfrak{W}) = \inf_{t \in \theta} C_S(\mathfrak{W}_t).$$

Now, additionally let us assume that each V_s is a degraded version of every W_t for $s \in \mathcal{S}$ and $t \in \theta$. Then as shown in lemma 3.9 $I(X; Y_t | Z_s)$ is a concave function with respect to the input distribution $p_X = p$. In particular this still holds for $\min_{s \in \mathcal{S}} I(X; Y_t | Z_s)$. Now because the random variables X, Y_t, Z_s form a Markov chain for all $t \in \theta$ and $s \in \mathcal{S}$ and

$$\min_{s \in \mathcal{S}} I(X; Y_t | Z_s) = I(X, Y_t) - \max_{s \in \mathcal{S}} I(X, Z_s),$$

for any $t \in \theta$ we get the upper bound on the secrecy rate as the secrecy capacity of a single channel W_t with S channels to the eavesdropper. Then we can conclude

Proposition 3.13: The secrecy capacity of the channel where only the channel states to the legitimate receiver are known and the channels to the eavesdropper are degraded versions of those to the legitimate receiver is given by

$$C_{S,CSI_t}(\mathfrak{W}) = \min_{t \in \theta} \max_{p \in \mathcal{P}(A)} (I(p, W_t) - \max_{s \in \mathcal{S}} I(p, V_s)).$$

E. A compound wiretap channel with $C_S = C_{S,CSI}$

Let $\mathfrak{W} := \{W_t, V_s : t = 1, \dots, T, s = 1, \dots, S\}$ with $S \neq T$ and the pair (t, s) unknown to both the transmitter and the legitimate receiver. In addition let us assume that

$$\exists \hat{t} \in \theta \forall t \in \theta \exists U_t : W_{\hat{t}} = U_t W_t, \quad (47)$$

which means that $W_{\hat{t}}$ is a degraded version of all channel W_t with $t \neq \hat{t}$. We further assume that

$$\exists \hat{s} \in \mathcal{S} \forall s \in \mathcal{S} \exists \hat{U}_s : V_s = \hat{U}_s V_{\hat{s}}, \quad (48)$$

which means that all V_s with $s \neq \hat{s}$ are degraded versions of $V_{\hat{s}}$. Then we can show that the capacity of this channel equals the capacity of the same channel with CSI at the transmitter, e.g.

$$C_S(\mathfrak{W}) = C_{S,CSI}(\mathfrak{W}).$$

First, by Theorem 3.6 it holds that

$$C_S(\mathfrak{W}) \geq \max_{p \in \mathcal{P}(A)} \min_{(t,s)} (I(p, W_t) - I(p, V_s)). \quad (49)$$

Now let

$$p^* = \arg \max_{p \in \mathcal{P}(A)} (I(p, W_{\hat{t}}) - I(p, V_{\hat{s}}))$$

the input distribution that achieves capacity for the single wiretap channel $(W_{\hat{t}}, V_{\hat{s}})$. Because the capacity of the compound

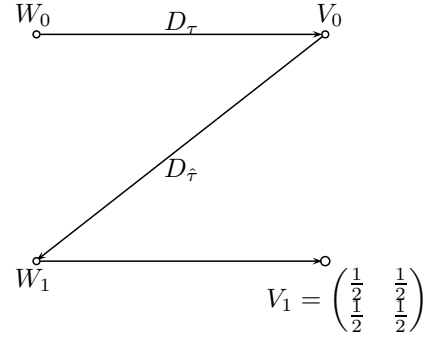


Fig. 1. Compound wiretap channel $\mathfrak{W} = \{(W_t, V_t) : t = 0, 1\}$ of Ex. 1

wiretap channel \mathfrak{W} is less than or equal the capacity of each single channel we obtain

$$\begin{aligned} C_{S,CSI}(\mathfrak{W}) &\leq I(p^*, W_{\hat{t}}) - I(p^*, V_{\hat{s}}) = C_S(W_{\hat{t}}, V_{\hat{s}}) \\ &\leq I(p^*, U_t W_t) - I(p^*, \hat{U}_s V_{\hat{s}}) \\ &\leq I(p^*, W_t) - I(p^*, V_s) \end{aligned} \quad (50)$$

for all $(s, t) \in \mathcal{S} \times \theta$ because of (47), (48). Then by the last inequality it follows that

$$\begin{aligned} I(p^*, W_{\hat{t}}) - I(p^*, V_{\hat{s}}) &= \min_{(s,t)} (I(p^*, W_t) - I(p^*, V_s)) \\ &\leq \max_{p \in \mathcal{P}(A)} \min_{(s,t)} (I(p, W_t) - I(p, V_s)) \end{aligned}$$

Now taking into account (49) and (50) we end in

$$C_{S,CSI}(\mathfrak{W}) \leq C_S(\mathfrak{W})$$

and therewith for this channel the lower bound of the capacity without CSI matches the capacity of the compound wiretap channel with CSI.

IV. EXAMPLES

In this section we provide some examples which display some striking features of compound wiretap channels as opposed to the usual compound channels. Our first example shows clearly that for compound wiretap channels with CSI at the transmitter the strategy of sending both the message and the randomisation parameter does not work. The second one demonstrates that even in the case where the sets of channels to the legitimate receiver and the eavesdropper both are convex, we can have

$$C_{S,CSI}(\mathfrak{W}) > 0 \text{ and } C_S(\mathfrak{W}) = 0,$$

as opposed to the case of the usual compound channel where the Minimax-Theorem applies.

In the following we use some simple facts which we state here without proof.

Fact 1. The binary entropy function

$$h(x) := -x \log x - (1-x) \log(1-x), \quad x \in [0, 1],$$

is strictly increasing on $[0, \frac{1}{2}]$.

Fact 2. Let $\eta \in [0, 1]$ and set

$$D_\eta := \begin{pmatrix} 1-\eta & \eta \\ \eta & 1-\eta \end{pmatrix}.$$

Then for every $\tau, \tau' \in [0, 1]$ it follows that

$$D_\tau D_{\tau'} = D_{\tau+\tau'-2\tau\tau'}.$$

Moreover, if $\tau, \tau' \in (0, \frac{1}{2})$ then

$$\tau + \tau' - 2\tau\tau' \in (0, \frac{1}{2})$$

and

$$\tau + \tau' - 2\tau\tau' > \tau, \tau'.$$

Fact 3. For $\tau, t \in [0, 1]$

$$(1-t)D_0 + tD_\tau = D_{t\tau}.$$

A. Example 1

Consider a compound wiretap channel $\mathfrak{W} = \{(W_t, V_t) : t \in [0, 1]\}$ in the case of CSI at the transmitter. First we define the channels to the legitimate receiver and to the eavesdropper for $t = 0$ by

$$W_0 = D_\eta, \quad \eta \in [0, \frac{1}{2}), \quad \eta \approx 0,$$

$$V_0 := D_\tau W_0, \quad \tau \in [0, \frac{1}{2}), \quad \tau \approx 0,$$

and for $t = 1$, $\hat{\tau} \in (0, 1/2]$

$$W_1 := D_{\hat{\tau}} V_0 = D_{\hat{\tau}} D_\tau W_0,$$

$$V_1 := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Hence V_0 and W_1 are degraded versions of W_0 and

$$I(p, V_1) = 0, \quad \forall p \in \mathcal{P}(A)$$

by definition of V_1 . Now for every $p \in \mathcal{P}(A)$ we can choose τ small enough, that

$$I(p, W_0) - I(p, V_0) < I(p, W_1).$$

Now with

$$p_0 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

$\nu > 0$ and because we have CSI at the transmitter we have by the defining equations (13) and (14)

$$\begin{aligned} J_n &= 2^{n[I(p_0, W_0) - I(p_0, V_0)] - \nu} \\ L_{n,0} &= 2^{n[I(p_0, V_0) + \frac{\nu}{4}]} \end{aligned}$$

such that we obtain

$$J_n L_{n,0} = 2^{n[I(p_0, W_0) - \frac{3\nu}{4}]}$$

But for $\hat{\tau}$ close to $1/2$ it holds then that

$$\begin{aligned} I(p_0, W_0) - \frac{3\nu}{4} &> I(p_0, W_1) \\ &= \max_{p \in \mathcal{P}(A)} I(p, W_1) = C_{CSI}\{W_0, W_1\}, \end{aligned}$$

where $C_{CSI}\{W_0, W_1\}$ is the capacity of a compound channel with CSI at the transmitter. Hence we have shown, that we can achieve reliable transmission of the message $j \in [J_n]$, but identifying both the message and the randomizing indices is not possible for all pairs $j \in [J_n]$ and $l \in [L_{n,t}]$. This is in contrast to the case where we have only one channel to both the legitimate receiver and the eavesdropper (cf. [7], [4]).

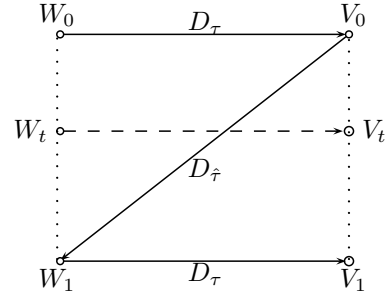


Fig. 2. Compound wiretap channel $\mathfrak{W} := \{(W_t, V_t) : t \in [0, 1]\}$ of Ex. 2

B. Example 2

Now, for $\eta, \tau \in (0, \frac{1}{2})$ we set

$$W_0 = D_\eta, \quad V_0 := D_\tau W_0 = D_{\eta+\tau-2\eta\tau}$$

and

$$W_1 := D_\tau V_0 = D_{2\tau-2\tau^2} W_0, \quad V_1 := D_\tau W_1.$$

Notice that V_0 is a degraded version of W_0 , W_1 of V_0 , and V_1 of W_1 . Next we define for $t \in [0, 1]$

$$\begin{aligned} W_t &:= (1-t)W_0 + tW_1 \\ &= [(1-t)D_0 + tD_{2\tau-2\tau^2}]W_0 \end{aligned} \quad (51)$$

and

$$\begin{aligned} V_t &:= (1-t)V_0 + tV_1 \\ &= D_\tau [(1-t)D_0 + tD_{2\tau-2\tau^2}]W_0 \\ &= D_\tau W_t \end{aligned} \quad (52)$$

By the definition, the set of channels to the legitimate receiver $\{W_t\}$ and the set of channels to the eavesdropper $\{V_t\}$ both are convex. Nevertheless we will show now, that for the compound wiretap channel $\mathfrak{W} := \{(W_t, V_t) : t \in [0, 1]\}$ we have

$$C_{S,CSI}(\mathfrak{W}) > 0, \quad C_S(\mathfrak{W}) = 0.$$

To this end, note that by (51), fact 3, and fact 2 we have

$$W_t = D_{t(2\tau-2\tau^2)} D_\eta = D_{f(t,\eta,\tau)}$$

with

$$f(t, \eta, \tau) := \eta + t(2\tau - 2\tau^2) - 2\eta t(2\tau - 2\tau^2) \in (0, \frac{1}{2}). \quad (53)$$

Similarly from (52) and fact 2 we obtain

$$\begin{aligned} V_t &= D_\tau D_{f(t,\eta,\tau)} \\ &= D_{\tau+f(t,\eta,\tau)-2\tau f(t,\eta,\tau)} \end{aligned}$$

Additionally from (53) and fact 2 we get

$$\tau + f(t, \eta, \tau) - 2\tau f(t, \eta, \tau) \in (0, \frac{1}{2})$$

and

$$\tau + f(t, \eta, \tau) - 2\tau f(t, \eta, \tau) > f(t, \eta, \tau). \quad (54)$$

Taking $p = (1/2, 1/2)$ we obtain for every $t \in [0, 1]$

$$\begin{aligned} I(p, W_t) - I(p, V_t) \\ = h(\tau + f(t, \eta, \tau) - 2\tau f(t, \eta, \tau)) - h(f(t, \eta, \tau)) > 0 \end{aligned}$$

where the last inequality follows from fact 1 and (54). Thus we have shown that

$$C_{S,CSI}(\mathfrak{W}) > 0$$

holds by Theorem 3.5.

In order to show that $C_S(\mathfrak{W}) = 0$, we have to employ our multiletter converse in the case of no CSI, Proposition 3.8. First, a simple algebra shows that for any $t \in [0, 1]$

$$V_t = ((1-t)D_0 + tD_{2\tau-2\tau^2})V_0$$

by (52) and thus each V_t is a degraded version of V_0 . Then the data processing inequality implies that for each $n \in \mathbb{N}$

$$\max_{t \in [0,1]} I(p, V_t^{\otimes n}) = I(p, V_0^{\otimes n}) \quad (55)$$

for all $p \in \mathcal{P}(\{0, 1\}^n)$.

On the other hand, since $W_1 = D_\tau V_0$ we obtain by data processing inequality and (55) for all $n \in \mathbb{N}$

$$I(p, W_1^{\otimes n}) - \max_{t \in [0,1]} I(p, V_t^{\otimes n}) = I(p, W_1^{\otimes n}) - I(p, V_0^{\otimes n}) \leq 0,$$

for all $p \in \mathcal{P}(\{0, 1\}^n)$. Then Proposition 3.8 implies that

$$C_S(\mathfrak{W}) = 0$$

as desired.

ACKNOWLEDGMENT

Support by the Deutsche Forschungsgemeinschaft (DFG) via projects BO 1734/16-1, BO 1734/20-1, and by the Bundesministerium für Bildung und Forschung (BMBF) via grant 01BQ1050 is gratefully acknowledged.

REFERENCES

- [1] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP Journal on Wireless Communications and Networking*, 2008.
- [2] M. Bloch and J.N. Laneman, "On the secrecy capacity of arbitrary wiretap channel," *Forty-Sixth Annual Allerton Conference, Allerton House, Illinois, USA*, Sep. 2008.
- [3] A. Wyner, "The wire-tap channel," *The Bell System Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
- [4] I. Csiszar, "Almost independence and secrecy capacity," *Problems of Information Transmission*, vol. 32, no. 1, pp. 40–47, 1996.
- [5] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Advances in Cryptology-Eurocrypt 2000, Lecture Notes in Computer Science*, vol. 1807, pp. 351–368, 2000.
- [6] N. Cai, A. Winter, and R. Yeung, "Quantum privacy and quantum wiretap channel," *Problems of Information Transmission*, vol. 40, no. 4, pp. 318–336, 2004.
- [7] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, no. 51, pp. 44–55, January 2005.
- [8] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, pp. 1229–1241, 1959.
- [9] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 15–29, 1989.
- [10] R. Ahlswede, "General theory of information transfer: updated," *General Theory of Information Transfer and Combinatorics, Special Issue of Discrete Applied Mathematics*, vol. 156, no. 9, pp. 1348–1388, 2008.
- [11] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, 1981.
- [12] R. F. Wyrembelski, I. Bjelaković, T. J. Oechtering, and H. Boche, "Optimal coding strategies for bidirectional broadcast channels under channel uncertainty," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2984–2994, October 2010.

- [13] M. Fekete, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Mathematische Zeitschrift*, vol. 17, pp. 228–249, 1923.
- [14] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography-part I: Secret sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.